Optimized Gaussian Process Regression for Prediction of Oil and Gas Pipelines Defect Length

Huda Aldosari[1], Raafat alfouly[2] and Reda Ammar[3]

[1,3] University of Connecticut,
[2] University of Rhode Island
[1,3]{huda.aldosari,reda.ammar@uconn.edu}
[2] relfouly@ric.edu

Abstract

Magnetic flux leakage (MFL) signals are used to estimate the scale and form of faults caused by the decaying metal used to build oil and gas pipelines. These faults, such as rust, can have catastrophic consequences if left undetected and improperly treated, both in terms of environmental damage and loss of life, as well as millions of dollars in maintenance costs for the stakeholders. Machine learning algorithms have proven their ability to solve the problem by correctly recognizing and calculating the scale and form of certain defects. The nonparametric and Bayesian approach to regression known as Gaussian process regression (GPR) is gaining popularity in machine learning. The optimization of GPR was carried out in this report using noisy and noiseless MFL signal measurements. The tune-able hyper-parameters were subjected to GPR optimization. Root mean square error (RMSE) error was used to calculate the output. In this research, the Quasi-Newton Method (QNM), an automated methodology for optimizing nonparametric regression analysis, was used to refine the GPR model. The optimization results are then compared to GPR analysis with default parameters, and it has been shown that QNM effectively optimizes the GPR while producing lower RMSE scores on all datasets. The ideal inferred parameter set can be used to train the GPR model for better output outcomes in determining oil and gas pipeline defects.

1 Introduction

Energy business involves the vital part of the transportation of gas and oil from the extraction sites of remote areas to the consumer. The means of this transportation industry are of various kinds (e.g., trucks, trains, and ships) for these products, but the safest one is the pipelines. One such example is Canada where pipelines are used as a transportation medium for 97% of the natural gas as well as the production of crude [1].

There are, however, certain drawbacks of these pipelines of oil and gas. They, for example, use delicate pathways that cross through nearby urban areas, or from fragile ecosystems which can increase the danger associated with any transportation failure or spillage of fuel spillage. These pipelines are comprised of many components and are laid over a longer stretch of distances increasing their failure potential. Due to their system's complexity, it is very difficult to inspect them, requiring extremely sophisticated technology. Generally, this inspection is comprised of an analysis of the scans from the walls of pipelines to recognize patterns for defect detection.

The magnetic flux leakage (MFL) inspection techniques are useful for acquiring high-resolution images and/or signals for detecting any anomalies and indications of defects in the interior wall of the pipeline so they can be duly reported [2]. This evaluation must be routinely done for integrity and to avoid any catastrophic failure or pipeline leak resulting in unfavorable effects from environmental damage. Due to the outrageous expense involved in shutting off a pipeline when the inspection is being done, autonomous operational devices called pipeline inspection gauges (PIGs) are designed which propel by the normal flow in the transport pipeline [3]. Another name for PIGs is "in-line-inspection (ILI)" tools. This ILI tool has the function to magnetize the wall of the pipeline during its travels down along the pipe. To measure the intensity of magnetic flux along the wall of this pipe for a localized leakage, coil sensors or Hall effect are used. Irregularities may occur in the magnetic field due to the defects present in the pipeline wall, for whose detection a regular sequential arrangement round the inner circumference lining of the wall of the pipe, these sensors are placed in the form of an array. So MFL testing is done based on the magnetic field detection from a pipe wall leakage in the subsurface flaws and the vicinity of the surface [4-7].

This stored data of the magnetic images is later analyzed offline. The pipeline operators have more interest in the degree and the location of different defects that may excessively affect the operation and the integrity of this pipeline. Some such defects are faulty welds, fatigue, deformation, buckles, hairline cracks, delamination, dents, and corrosion. The pipeline operator uses these results for determining the priority from replacement and repair. Typically, nondestructive testing (NDT) technicians are the human operators who do the image analysis part manually in the procedure of inspection which due to its nature is thus error-prone and inherently slow. To automate this inspection process, the industry is highly motivated for using such methods as machine learning. When identification of these defects has been done, another vital problem arises, which involves the assessment of the severity or size of this defect (sizing) [7]. For determining the safety of a pipe and for accurate calculation of operating pressure (MAOP) the use of 'estimated defect depths' is made for the flowing gas or oil through the pipeline [8]. There are various previous studies done for the use of the MFL technique in sizing and defect detection [9,10]. Machine learning has previously been used for this purpose, which does the evaluation of multilayer perceptron (MLP) use in weld joints of the pipelines for the pattern recognition of these MFL signals [11]. Approaches such as Inverse modeling are used for the defects in the shape construction [12]. An iterative method of inversion was proposed by Joshi et al. (2006) which made use of a "radial basis function neural network (RBFN)" and a multiresolution wavelet transforms for predicting the 3-D version of the defect in geometry by using measurements taken through MFL [13].

A domain adaptive finite impulse response-based wavelet transforms filtration framework to eliminate the seamless pipe noise from MFL signals was proposed by Afzal et al. (2002) and Han et al. (2006) [14,15]. The research articles [15] and [16] provide approaches that are wavelet-based for this problem in the case of both denoising and classifying. The finite impulse response filter is used in [17] in MFL inspection for presenting a method that is adaptive for the equalization of the channel (for compensating the sensors' mismatch). The natural gas pipelines for transmission are inspected in MFL by a model-based evaluation technique called the probability of detection (POD) in [18].

It is a challenging task to accurately estimate the likelihood of failure due to the involvement of larger uncertainties in the associated parameters of the prediction models for burst failure. Since the gas and oil pipelines are complex and typically larger due to the presence of hundreds of segments of pipelines, it renders the physical probabilistic approach to be computationally intensive in terms of

application. It is not often feasible for the decision-making of maintenance to carry out an analysis of reliability for evaluation of every single pipeline present in the complex transmission network of natural gas and oil. A significant effort is made by the utilities that are responsible for the maintenance of pipelines for developing approaches that are sustainable, cost-efficient, and of higher quality. There is a need for a more robust technique to overcome the lack of computational efficiency in the existing approaches for predicting the risk in ratings and pipe failure in handling large complex networks of these pipelines. In replacement, a new approach that is computationally efficient and has the potential of risk determination for all the individual segments in a complex system of the network without distinctly executing exhaustive analyses that are physically based. In a broad variety of areas, machine learning approaches are used, such as automatic diagnosis processes, credit card fraud identification, stock market monitoring, biomedical signal processing, and text recognition, autonomous systems, etc. [19-22]. Traditional probabilistic physical analysis techniques, for the estimation of oil and gas pipeline failure estimation, are computationally expensive. Thus, alternatively, machine learning-based data-driven approaches to predict the oil and gas defect sizes may help to overcome the extensive computational requirements.

Machine learning is a modern research dimension that is currently in the phase of active development. Several algorithms have transpired in the last decade, such as k-nearest neighbors, support vector machine, and random forest, etc. [19]. For predicting the probable failure of pipes with active defects due to corrosion Anghel et al. (2009) developed an innovative kind of vector machine that is maximally supportive [23]. A classification reliability procedure was developed by the author that initiated a connection between artificial intelligence (AI) and the methods of reliability. The following approach has provided a simpler and propitious alternate method to prioritize these maintenance works. Huang and Burton (2019) used the applications of machine learning for classifying the failure modes of the concrete frames that are reinforced with that of the infill walls [24]. Winkler et al. (2018) used the decision-tree-based algorithms of machine learning for predicting the failures of water pipes [25]. For improving the model accuracy, boosting techniques and Bootstrap aggregation are utilized. A confusion method evaluates the performance of the model. Mangalathu and Jeon identified the modes of failure of the bridge columns that are reinforced circularly by utilizing the various algorithms of machine learning [26]. The authors did a comparison, by using an experimental dataset, of the efficiency of several 6 different algorithms of machine learning. The artificial neural networks among the six algorithms have imparted the best efficiency. Kiani et al. have studied the applications of the algorithms of machine learning for deriving the fragility curves [27]. Classification of masonry walls by using quite a limited dataset was done by Siam et al. with the help of a clustering algorithm.

A data-driven technique was applied by Mangalathu et al. for the recognition of the modes of failure of the concrete shear walls, this was where these authors had used a variety of different models for boosting apart from other techniques of machine learning [28]. For the recognition of the failure modes, ten different input parameters are used for capturing the reinforcement, geometry, and material characteristics. The algorithm's efficiency in machine learning algorithms was established on the recall, global accuracy, and precision basis. This study has shown a positive result for the mode of failure for classifying which can be further applied over other various kinds of structures or infrastructures. Studies of Qi and Zhu, Yu et al., Zhang et al., and Jeon et al., gave other applications that are notable in machine learning [29-31]. Machine learning applications can solve complex problems even in the absence of explicit mechanical analysis, although, for determining the feasibility of various applications regarding the algorithms of machine learning, limited research has been done for the analysis of burst failure risk of the pipeline. These future pipe failure predictions might help the decision-makers in taking suitable intervention measures for avoiding any potential consequences of the burst failure.

Even though several machine learning algorithms exist and can predict the dimensioning of oil and gas defects, findings have shown that GPR outperforms other ML algorithms. In this paper, we extend the previous work by examining the best possible parameters for improving the efficiency of GPR and establishing high-performance machine learning strategies for pipeline fault detection [32,33]. The aspect of this study was to consider success patterns, with different sets of parameters determined by

measuring the performance of each mixture. A pre-recorded dataset of noisy and noiseless MFL signals was used in this analysis [34]. To reduce data dimensionality, indicative and distinguishing characteristics were first extracted from MFL signals; this, in turn, resulted in speeding up the learning process and improving the performance of the current system in terms of estimation accuracy. Statistical methods and polynomial series were used to extract advanced features from four MFL datasets, yielding a total of 33 features. The data was arranged as follows: 80% of the time is spent on teaching, and 20% on evaluation. The extracted features were fed into a Gaussian process regression (GPR) model. If correctly calibrated before training, the GPR may provide very satisfactory results for the detection of defects and the measurement of real defects with odd geometric arrangements.

The following organization has been used in this paper; In Section 2, we provide the details about the utilized MFL signals dataset and adopted methodology. Section 3 consists of the experimental results for the real MFL data, and conclusions are in Section 4.

2 Methodology

2.1 Dataset

Nondestructive Inspection (NDE) methods are frequently used to screen pipelines for potential defects. A magnetic sensor attached to an analysis computer is one such NDE instrument that is routinely sent to test oil and gas supply tubing. Magnetic sensors, spaced every 3 mm along the pipeline's circumference, are used to calculate MFL signals. This study used a pre-recorded MFL dataset of noiseless and noisy MFL signals at 3dB, 5dB, and 10dB signal-to-noise ratio (SNR) [34]. Since the amount of MFL data is too large, feature extraction techniques have been used to reduce the data's feature space. Using any of the derived functions, on the other hand, does not always result in stronger reliability test outcomes. The most important characteristics are thus selected and fed into the identifying and sizing units.

2.2 Feature Extraction

Feature extraction aims to reduce the size of the MFL data. The MFL signal is expressed in the axial, radial, and tangential dimensions. For each component, statistical and polynomial series are used as feature extraction methods. From original MFL signals, statistical characteristics of the integral normalized signal (INS), mean average (MA), maximum magnitude (MM), standard deviation (STD), and peak-to-peak distance (PPD) were derived, as well as polynomial rank 3, 6, and 6 series corresponding to the axial, radial, and tangential components, respectively. Polynomial coefficients make up the input characteristics, in addition to the five statistical characteristics. There are a total of 33 functions because of this.

2.3 Feature Reduction

Many researchers have shown that individuals with different attributes may have differing degrees of discrimination capacity. In certain cases, using the whole collection of retained features in the training phase results in poor defect prediction accuracy. Certain attributes have the potential to be harmful. As a result, determining the main attributes that make the ML model perfect is a routine practice. Following that, analyses are carried out to assess the suitability of each element for the defect length prediction mission. The most appropriate features that yield the most effective defect convergence rate are then found using neighborhood component analysis (NCA) as an input function pattern for the ML model. The features with a weight close to 0 make little difference in calculating defect sizes. Furthermore, for assigning weights for the retained functions, a principal component analysis-based weight correlation technique is used. The highest-weighted attributes are then used to

train the model. From the axial, radial, and tangential elements, nine characteristics were chosen based on their weights: INS and PPD; INS, PPD, and MM; and INS, PPD, MM, and STD from x, y, and z components, respectively.

2.4 Gaussian Process Regression

Gaussian process regression (GPR) is the nonparametric and a Bayesian approach to the regression that is currently making waves in the field of machine learning. There are several benefits of the GPR, which works well on considerably small datasets and can provide predictions with their uncertainty measurements. In GPR, firstly, a Gaussian process prior is assumed, can then be made specific using a covariance function and a mean function. A Gaussian process is particularly similar to an infinite-dimensional multivariate Gaussian distribution, where any dataset collection of the labels is distributed as joint Gaussian. Within this GP prior, prior knowledge can be incorporated about the space of the functions through the selection of the covariance and the mean functions. Gaussian noise that is identically distributed can also be incorporated independently with ease, to the labels by the summation of the label distribution and the noise distribution. A GPR algorithm predicts the outcome of Xnew given the new input vector Ynew and the training data. A linear regression model is

$$X = Y^T \beta + \varepsilon \qquad\qquad (1)$$

The data estimate the error variance $\varepsilon \sim N(0,\sigma^2)$ and the coefficients. GPR models describe the response by using latent variables from a Gaussian process and explicit basis functions, G. The latent variable covariance function measures response smoothness, while basis functions project input x onto a p-dimensional feature space.

During the selection of the model, the mean function form and the covariance kernel function in the GP is chosen priorly. The function of mean is generally constant, either it is zero or it is the mean of the training dataset. The function of covariance kernel has many options: if the kernel properties are followed, it may have many forms. Some of the common kernel functions comprise linear, constant, Matern kernel, square exponential, and multiple kernels compositions. Other choices for the GP model can now be specified after the kernel function has been specified. For example, the variance of noise on the labels is taken as alpha, and training data mean or the zero is referred to as the constant mean function.

By maximizing the likelihood of the log marginal of training data, the hyperparameters of the function of the covariance kernel can be tuned. Generally, a gradient-based optimizer is used for increasing the efficiency; if it is not specified above, then we use the default optimizer. The optimizer is restarted multiple times with varying initializations owing to the unlikeliness of the convexity of the log marginal. The training data and test data samples are filtered out of posterior distribution, to estimate predictive posterior distribution. As normal distribution can be explained by mean and covariance thus, we opt for the Gaussian processing prior. The following models are estimated to fit a GPR model from the data:

- Covariance function parameterized in terms of kernel parameters.

- Noise variance, and

- Coefficient vector of fixed basis functions.

The 'Sigma' consists of the initial value for noise standard deviation. When the optimization is ongoing, the GPR creates a vector of values from unconstrained initial parameters by utilizing the values obtained initially for the standard deviation of noise and the kernel parameters.

2.5　Quasi-Newton Method Optimization

Quasi-Newton Methods (QNMs) are generally a class of optimization methods that are used in Non-Linear Programming when full Newton's Methods are either too slow or not easy to use, such as finding the global minimum of a certain function which can be differentiated two times. Quasi-Newton Methods have definite benefits as compared to the full Newton's Method for the problems of non-linearity that are complex and expansive. These methods are thus imperfect and may have certain flaws which depend on the problem and the type of Quasi-Newton Method used. Despite this, apart from few simple problems, Quasi-Newton Methods are technically worthy. QNM are algorithms for discovering local maxima and minima and use Newton's approach to determine a function's stationary point. The first and second derivatives are used to determine the stationary point in Newton's technique. In higher dimensions, Newton's technique utilizes the gradient and Hessian matrix of second derivatives.

In this study, the aim was to optimize GPR analysis for the prediction of oil and gas pipeline defects. For the said purpose, QNM optimization was used to find the best model parameters. In our previous paper, it was shown that GPR analysis has comparable results in comparison with other ML models. In this study, the results of optimized GPR analysis have been used with default GPR parameters.

3　Results

The dataset was divided in such a way that 80% of the data was used to train the GPR model and the remaining 20% to test the trained model. The results of the training set are presented only. Root mean square error (RMSE) has been used as a performance metric to calculate the performance of the model being practiced.

The study aimed to optimize GPR analysis to find the best possible parameters yielding the lowest RMSE value corresponding to each investigated dataset to predict the oil and gas pipeline defects. The GPR was first trained on default parameters and then it was optimized using QNM optimizer. GPR was trained on the parameters mentioned in table 1 before optimization. The resulting RMSE scores yielded from default parameters are summarized in table 2.

Parameter	Parameter Option
Method to estimate parameters	Exact Gaussian process regression
Explicit basis function	Constant
Noise standard deviation	2
Covariance function	Exponential
Method for computing inter-point distances	Fast
Active set selection method	Random selection
The method used to make predictions	Exact

Table 1: PERFORMANCE RESULTS OF VARIOUS TRAINING FUNCTIONS

To show that the QNM helps to identify the best possible parameters for the optimization of GPR analysis, the resulting RMSE values for each dataset are presented in table 2 corresponding to each dataset before and after optimization.

Dataset	Default	Optimized
Noiseless	0.2707	0.1371
3 dB Noise	0.3161	0.2821
5 dB Noise	0.3372	0.2909
10 dB Noise	0.4062	0.3933

Table 2: RMSE FOR DEFAULT AND OPTIMIZED GPR

From table 2, it can be seen that the QNM optimizer helps to improve the RMSE score for each dataset by finding the best parameters for GPR analysis. The optimizer yielded the parameters, as mentioned in table 3, after optimizing the GPR model corresponding to each dataset.

Parameter	Noiseless	0 dB	3 dB	10 dB
Method to estimate parameters	Fully independent conditional approximation	Fully independent conditional approximation	Fully independent conditional approximation	Fully independent conditional approximation
Explicit basis function	Linear	Linear	Linear	Pure Quadratic
Noise standard deviation	1.7	1.6	0.9	0.7
Covariance function	Matern	Matern	Matern	Exponential
Method for computing inter-point distances	Fast	Accurate	Accurate	Accurate
Active set selection method	Random selection	Differential entropy-based selection	Differential entropy-based selection	Sparse greedy matrix approximation
The method used to make predictions	Exact	Exact	Exact	Fully independent conditional approximation

Table 3: OPTIMIZED PARAMETERS FOR ALL DATASETS

Figure 1 shows the performance comparison of GPR analysis for both default parameters and after optimization. It can be graphically visualized that in all datasets the optimization results yield better predictions in comparison with default GPR analysis.
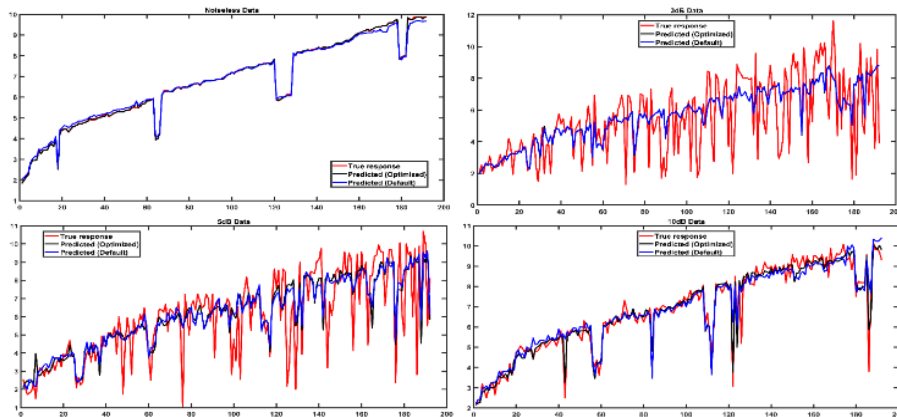

Figure 1: Performance comparison of default and optimized GPR analysis for all datasets

4   Conclusion

The basic aim of this study was to determine the optimum configurations for GPR to estimate the defect sizing using MFL signals. The performance of the GPR has been investigated in terms of various training hyper-parameters. MFL signals from noisy and noiseless datasets were used to train and test the GPR model. The performance results of optimum parameter configurations are compared with results yielded from default parameters and it has been shown that the presented results outperform the default configuration. The performance of the presented approach should be compared with optimization algorithms. The presented values of parameters are identified Quasi-Newton optimizers for better performance results, however, in the future other optimization techniques may also be investigated for automatic optimization of GPR. The inferred optimum parameter selection can be used to train the GPR model for enhanced performance results for the estimation of oil and gas pipeline defects.

References

[1]. Layouni, M., Tahar, S. and Hamdi, M.S., 2014, December. A survey on the application of neural networks in the safety assessment of oil and gas pipelines. In 2014 IEEE Symposium on Computational Intelligence for Engineering Solutions (CIES) (pp. 95-102). IEEE.

[2]. Pandey, M.D., 1998. Probabilistic models for condition assessment of oil and gas pipelines. Ndt & E International, 31(5), pp.349-358.

[3]. Kabir, G., Sadiq, R. and Tesfamariam, S., 2016. A fuzzy Bayesian belief network for safety assessment of oil and gas pipelines. Structure and Infrastructure Engineering, 12(8), pp.874-889.

[4]. V. E. Loskutov, A. F. Matvienko, B. V. Patramanskii, and V. E. Shcherbinin, "The magnetic method for in-tube nondestructive testing of gas and oil pipelines: The past and the present," Russ. J. Nondestruct. Test., vol. 42, no. 8, pp. 493–504, Aug. 2006.

[5]. R. C. Ireland and C. R. Torres, "Finite element modelling of a circumferential magnetiser," Sens. Actuators A (Phys.), vol. 129, no. 1–2, pp. 197–202, May 2006.

[6]. S. O'Connor, L. Clapham, and P. Wild, "Magnetic flux leakage inspection of tailor-welded blanks," Meas. Sci. Technol., vol. 13, no. 2, pp. 157–162, Feb. 2002

[7]. W. Lord, Nondestructive Testing Monographs and Tracts, Volume 3—Electromagnetic Methods of Nondestructive Testing. New York: Gordon and Breach, 1985.

[8]. Khodayari-Rostamabad, A., Reilly, J.P., Nikolova, N.K., Hare, J.R. and Pasha, S., 2009. Machine learning techniques for the analysis of magnetic flux leakage images in pipeline inspection. IEEE Transactions on magnetics, 45(8), pp.3073-3084.

[9]. Aldosari, H., Elfouly, R.S., Ammar, R. and Alsulami, M., 2020, March. New Monitori Architectures for underwater oil/Gas Pipeline using Hyper sensors. In CATA (pp. 307-316).

[10].   Aldosari, H., Elfouly, R.S., Ammar, R. and Alsulami, M., 2020, July. Performance of New Monitoring Architectures for Underwater Oil/Gas Pipeline using Hyper-Sensors. In 2016 IEEE international conference on cloud computing technology and science (CloudCom) (pp. 590-595). IEEE.

[11].   Khodayari-Rostamabad, A., Reilly, J.P., Nikolova, N.K., Hare, J.R. and Pasha, S., 2009. Machine learning techniques for the analysis of magnetic flux leakage images in pipeline inspection. IEEE Transactions on magnetics, 45(8), pp.3073-3084.

[12].   K. C. Hari, M. Nabi, and S. V. Kulkarni, "Improved FEM model for defect-shape construction from MFL signal by using genetic algorithm," IET Sci., Meas. Technol., vol. 1, no. 4, pp. 196–200, Jul. 2007.

[13].   A. Joshi, L. Udpa, S. Udpa, and A. Tamburrino, "Adaptive wavelets for characterizing magnetic flux leakage signals from pipeline inspection," IEEE Trans. Magn., vol. 42, no. 10, pp. 3168–3170, Oct. 2006.

[14].   M. Afzal and S. Udpa, "Advanced signal processing of magnetic flux leakage data obtained from seamless gas pipeline," NDT&E Int., vol. 35, no. 7, pp. 449–457, Oct. 2002.

[15].   W. Han and P. Que, "A modified wavelet transform domain adaptive FIR filtering algorithm for removing the SPN in the MFL data," Measurement, vol. 39, no. 7, pp. 621–627, Aug. 2006.

[16].   J. Tao, Q. Peiwen, C. Liang, and L. Liang, "Research on a recognition algorithm for offshore-pipeline defects during magnetic-flux inspection," Russ. J. Nondestruct. Test., vol. 41, no. 4, pp. 231–238, Apr. 2005.

[17].   Y. Zhang, Z. Ye, and X. Xu, "An adaptive method for channel equalization in MFL inspection," NDT&E Int., vol. 40, no. 2, pp. 127–139, Mar. 2007.

[18].   Z. Zeng, L. Xuan, Y. Sun, L. Udpa, and S. Udpa, "Probability of detection model for gas transmission pipeline inspection," Res. Nondestruct. Eval., vol. 15, no. 3, pp. 99–110, Jul./Sep. 2004.

[19].   Ashraf, H., Waris, A., Gilani, S.O., Kashif, A.S., Jamil, M., Jochumsen, M. and Niazi, I.K., 2021. Evaluation of windowing techniques for intramuscular EMG-based diagnostic, rehabilitative and assistive devices. Journal of Neural Engineering, 18(1), p.016017.

[20].   Ashraf, H., Waris, A., Jamil, M., Gilani, S.O., Niazi, I.K., Kamavuako, E.N. and Gilani, S.H.N., 2020. Determination of Optimum Segmentation Schemes for Pattern Recognition-Based Myoelectric Control: A Multi-Dataset Investigation. IEEE Access, 8, pp.90862-90877.

[21].   Asif, A.R., Waris, A., Gilani, S.O., Jamil, M., Ashraf, H., Shafique, M. and Niazi, I.K., 2020. Performance Evaluation of Convolutional Neural Network for Hand Gesture Recognition Using EMG. Sensors, 20(6), p.1642.

[22].   H. Ashraf, A. Waris, S. O. Gilani, M. U. Tariq and H. Alquhayz, "Threshold parameters selection for empirical mode decomposition-based emg signal denoising," Intelligent Automation & Soft Computing, vol. 27, no.3, pp. 799–815, 2021.

[23].   Anghel, C.I., 2009. Risk assessment for pipelines with active defects based on artificial intelligence methods. International journal of pressure vessels and piping, 86(7), pp.403-411.

[24].   Huang, H. and Burton, H.V., 2019. Classification of in-plane failure modes for reinforced concrete frames with infills using machine learning. Journal of Building Engineering, 25, p.100767.

[25].   Winkler, D., Haltmeier, M., Kleidorfer, M., Rauch, W. and Tscheikner-Gratl, F., 2018. Pipe failure modelling for water distribution networks using boosted decision trees. Structure and Infrastructure Engineering, 14(10), pp.1402-1411.

[26].   Mangalathu, S. and Jeon, J.S., 2018. Classification of failure mode and prediction of shear strength for reinforced concrete beam-column joints using machine learning techniques. Engineering Structures, 160, pp.85-94.

[27].   Kiani, J., Camp, C. and Pezeshk, S., 2019. On the application of machine learning techniques to derive seismic fragility curves. Computers & Structures, 218, pp.108-122.

[28].   Mangalathu, S., Hwang, S.H. and Jeon, J.S., 2020. Failure mode and effects analysis of RC members based on machine-learning-based SHapley Additive exPlanations (SHAP) approach. Engineering Structures, 219, p.110927.

[29].   Qi, F. and Zhu, A.X., 2011. Comparing three methods for modeling the uncertainty in knowledge discovery from area-class soil maps. Computers & Geosciences, 37(9), pp.1425-1436.

[30].   Yu, L., Porwal, A., Holden, E.J. and Dentith, M.C., 2012. Towards automatic lithological classification from remote sensing data using support vector machines. Computers & Geosciences, 45, pp.229-239.

[31].   Zhang, Y., Burton, H.V., Sun, H. and Shokrabadi, M., 2018. A machine learning framework for assessing post-earthquake structural safety. Structural safety, 72, pp.1-16.

[32].   H. Aldosari, R. Elfouly and R. Ammar, "Optimal Artificial Neural Network Model For Prediction of Oil and Gas Pipelines Defect Length," in 2020 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 2020 pp. 1457-1462. doi: 10.1109/CSCI51800.2020.00272

[33].   H. Aldosari, R. Elfouly and R. Ammar, "Evaluation of Machine Learning-Based Regression Techniques for Prediction of Oil and Gas Pipelines Defect," in 2020 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 2020 pp. 1452-1456.doi: 10.1109/CSCI51800.2020.00271

[34].   Layouni, M., Hamdi, M.S. and Tahar, S., 2017. Detection and sizing of metal-loss defects in oil and gas pipelines using pattern-adapted wavelets and machine learning. Applied Soft Computing, 52, pp.247-261.